

S.O.D.A.

Smart Osm Difference Analyzer



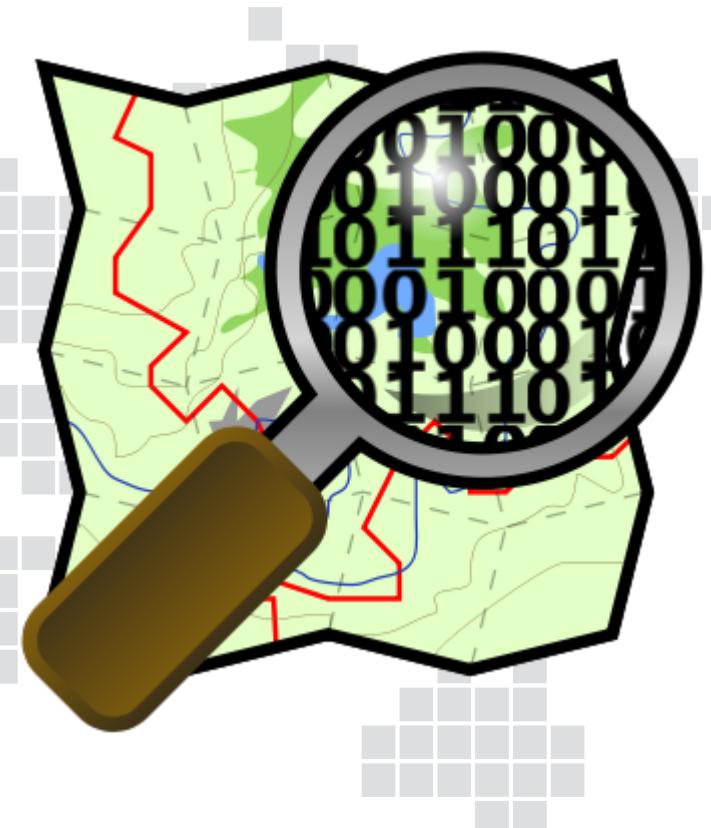
1. Introduction

- Problématique
- Outils existants

2. Soda

- Buts
- Principes
- Modules d'analyse
- Analyses existantes

3. Perspectives



• Contexte

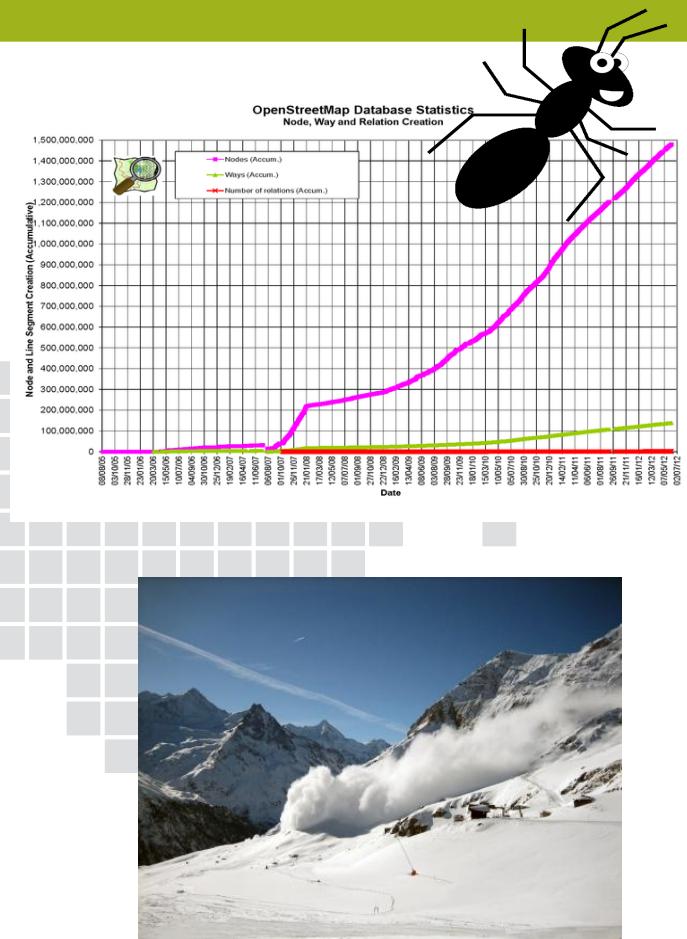
- Nombre croissants de contributeurs
- De plus en plus de données
- De plus en plus de modifications

• Problématique

- Comment “suivre” ce flux de données ?

• Concernés

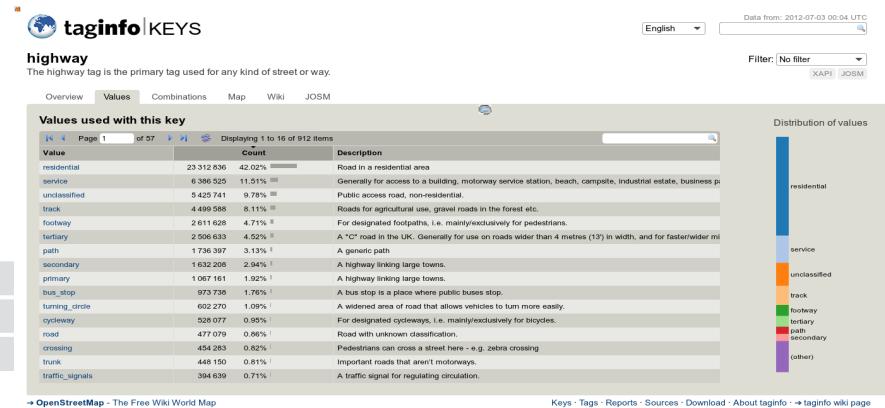
- Contributeurs
- Consommateurs de données
- Fournisseurs de données



Outils existants

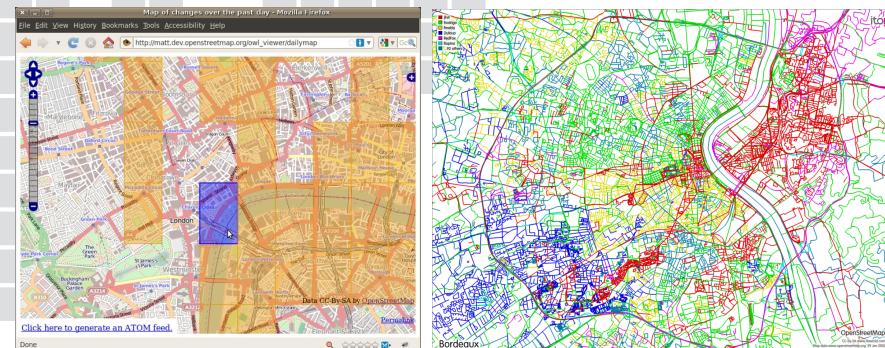
• Suivi des attributs

- TAG Info
<http://taginfo.osm.org>
- OSM Watch
<http://osm102.openstreetmap.fr/~zorglub/watch/>



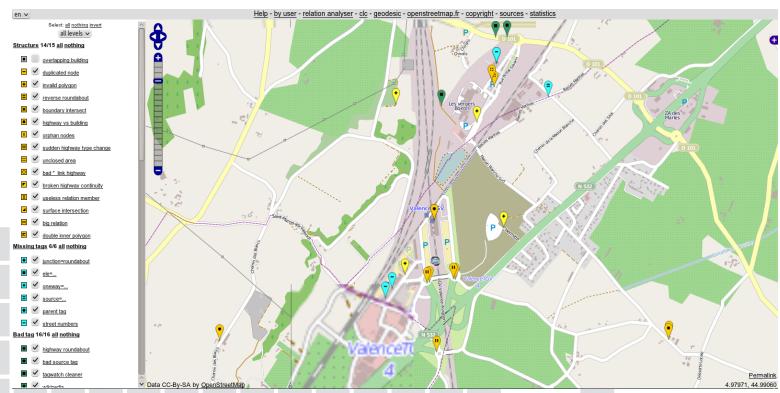
• Suivi des modifications

- OSM Mapper
<http://www.itoworld.com>
- OSM Watch list
<http://owl.apis.dev.openstreetmap.org/>
- Live OSM-Fr
<http://live.openstreetmap.fr/>



•Controle Qualité

- OSM Inspector
<http://tools.geofabrik.de/osmi>
 - Keep Right
<http://keepright.ipax.at>
 - Osmose
<http://osmose.openstreetmap.fr>



•Génération d'animations

- Osm-history tools
<https://github.com/MaZderMind/osm-history-splitter>



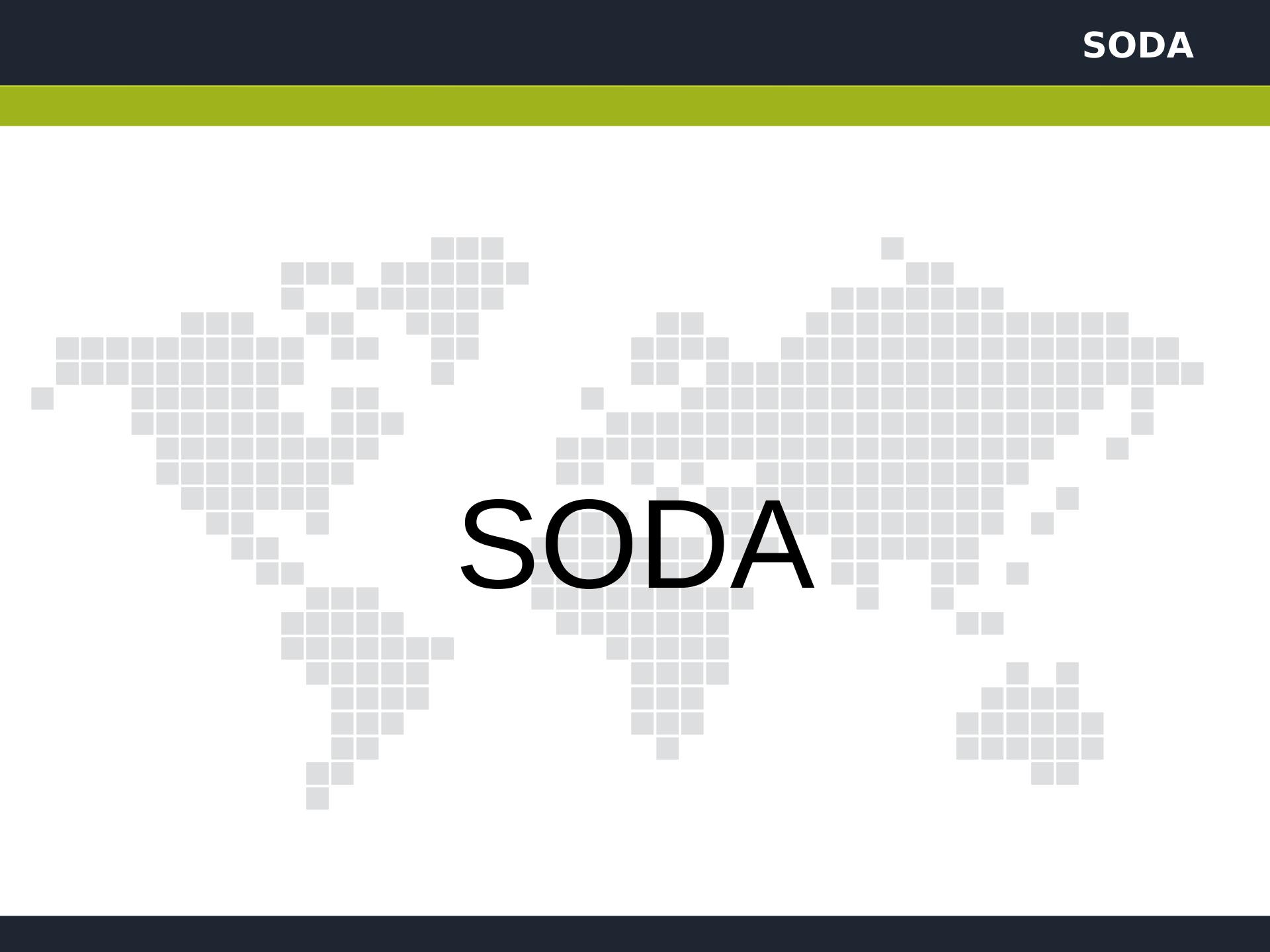
- Inconvénients

- Soit spécialisés soit généralistes mais “basiques”
- Peu personnalisables
- En ligne ou lourds et complexes à mettre en oeuvre
- Peu d'analyse dans le passé

- Solution

- Développer une autre approche

SODA

The background of the slide features a large, abstract pixelated pattern composed of numerous small gray squares. This pattern is dense and irregular, creating a sense of depth and texture. It is set against a white background and is partially obscured by the central text.

SODA

- But

- Permettre d'analyser les modifications d'OSM
 - En live ou à posteriori
 - Analyses simples ou avancées
 - Analyses personnalisables

- Cahier des charges

- Déploiement simple et rapide
- Configuration requise raisonnable
- Multi-plateforme
- Utilisation “simple”
- Fournir un framework aux développeurs

- Pour les utilisateurs

- Suivi thématique
- Génération d'alertes
- Bilan d'une cartopartie
- Surveillance de données sensibles
- Suivi d'import
- Détection de vandalisme
- Analyses statistiques
- Synchronisation avec des données externes

- Pour les développeurs

- Faciliter l'écriture d'analyses spécifiques via des services génériques

➤ Programme basé sur l'analyse des diffs OSM

- Avantages

- Service de base d'OSM
- Granularité variable
 - Minute, heures, jours
 - Planète, Europe, France
- Générées en live
- Historique disponible
 - Minutes monde: 27/09/2011
 - Heures monde: 19/11/2009
 - Jours monde: 28/11/2011

- Inconvénients

- Peu pratique pour les suppressions
- Pas d'historique complet
- Fragments de base

- Programme exécutable

- Capable de charger des bibliothèques de modules d'analyse
- Fournit des services génériques
 - Lecture des diffs, des fichiers OSM
 - Accès à l'API
 - Cache de données
 - Génération d'URL
 - Communication Interface Utilisateur ↔ modules

- Bibliothèques binaires

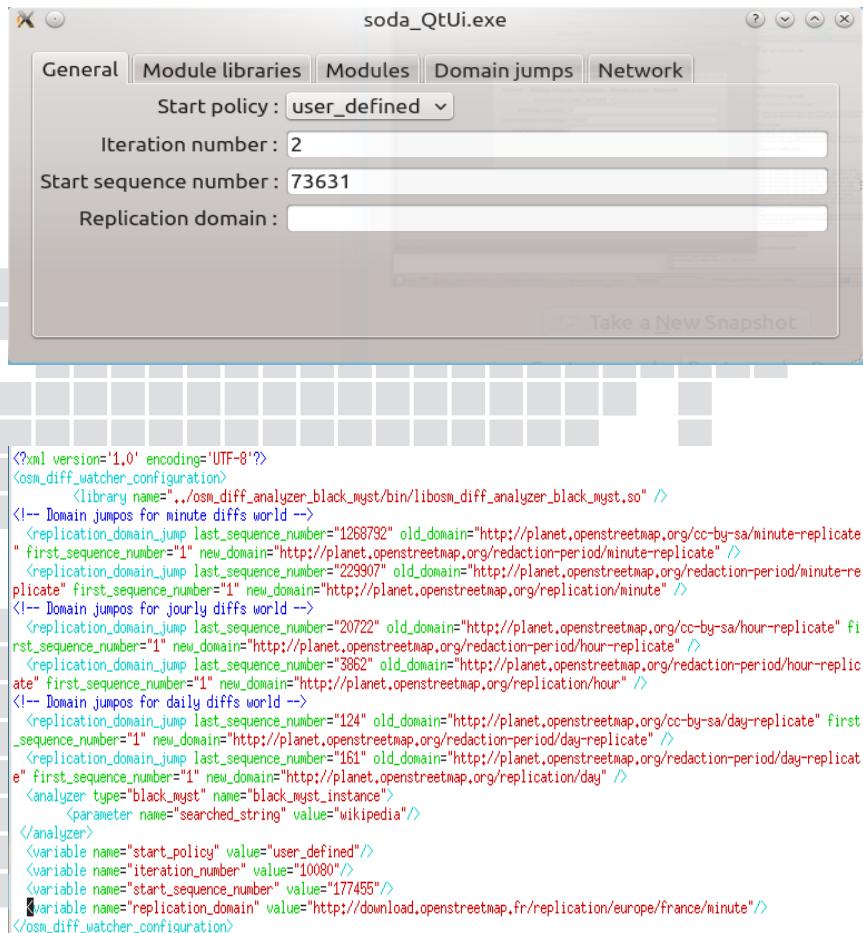
- Modules d'analyses spécifiques
- Utilisent les services génériques
- Génèrent des sorties
 - Rapport HTML, Widget graphique, messages IRC...

Configuration

- Via l'interface en mode graphique
- Via un fichier XML en mode batch

- Contenu

- Politique de démarrage
- Domaine de réPLICATION
- Emplacement des librairies de modules
- Configuration des modules
- Configuration de proxy
- Sauts entre domaines de réPLICATION

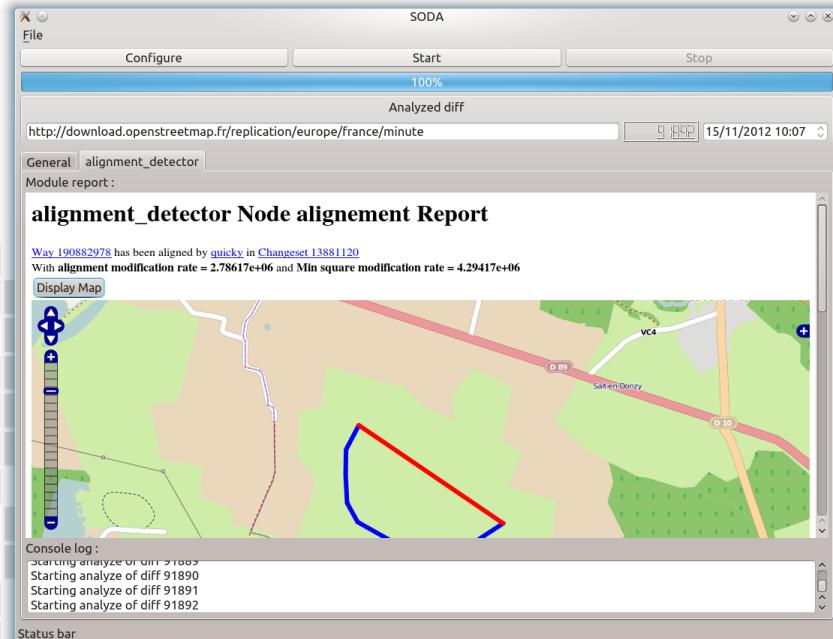


The screenshot shows a window titled "soda_QtUi.exe" with several tabs: General, Module libraries, Modules, Domain jumps, and Network. The General tab is selected, displaying configuration parameters: Start policy set to "user_defined", Iteration number set to 2, Start sequence number set to 73631, and Replication domain left empty. Below the window, a portion of an XML configuration file is visible:

```
<?xml version='1.0' encoding='UTF-8'?>
<osm_diff_watcher_configuration>
    <library name=".//osm_diff_analyzer_black_myst/bin/libosm_diff_analyzer_black_myst.so" />
    <!-- Domain jumps for minute diffs world -->
    <replication_domain_jump last_sequence_number="1268792" old_domain="http://planet.openstreetmap.org/cc-by-sa/minute-replicate"
        first_sequence_number="1" new_domain="http://planet.openstreetmap.org/redaction-period/minute-replicate" />
    <replication_domain_jump last_sequence_number="22907" old_domain="http://planet.openstreetmap.org/redaction-period/minute-replicate"
        first_sequence_number="1" new_domain="http://planet.openstreetmap.org/redaction-period/minute" />
    <!-- Domain jumps for hourly diffs world -->
    <replication_domain_jump last_sequence_number="20722" old_domain="http://planet.openstreetmap.org/cc-by-sa/hour-replicate"
        first_sequence_number="1" new_domain="http://planet.openstreetmap.org/redaction-period/hour-replicate" />
    <replication_domain_jump last_sequence_number="3862" old_domain="http://planet.openstreetmap.org/redaction-period/hour-replicate"
        first_sequence_number="1" new_domain="http://planet.openstreetmap.org/redaction-period/hour" />
    <!-- Domain jumps for daily diffs world -->
    <replication_domain_jump last_sequence_number="124" old_domain="http://planet.openstreetmap.org/cc-by-sa/day-replicate"
        first_sequence_number="1" new_domain="http://planet.openstreetmap.org/redaction-period/day-replicate" />
    <replication_domain_jump last_sequence_number="161" old_domain="http://planet.openstreetmap.org/redaction-period/day-replicate"
        first_sequence_number="1" new_domain="http://planet.openstreetmap.org/redaction-period/day-replicate" />
    <analyzer type="black_myst" name="black_myst_instance">
        <parameter name="searched_string" value="wikipedia"/>
    </analyzer>
    <variable name="start_policy" value="user_defined"/>
    <variable name="iteration_number" value="10080"/>
    <variable name="start_sequence_number" value="177465"/>
    <variable name="replication_domain" value="http://download.openstreetmap.fr/replication/europe/france/minute"/>
</osm_diff_watcher_configuration>
```

• Graphique

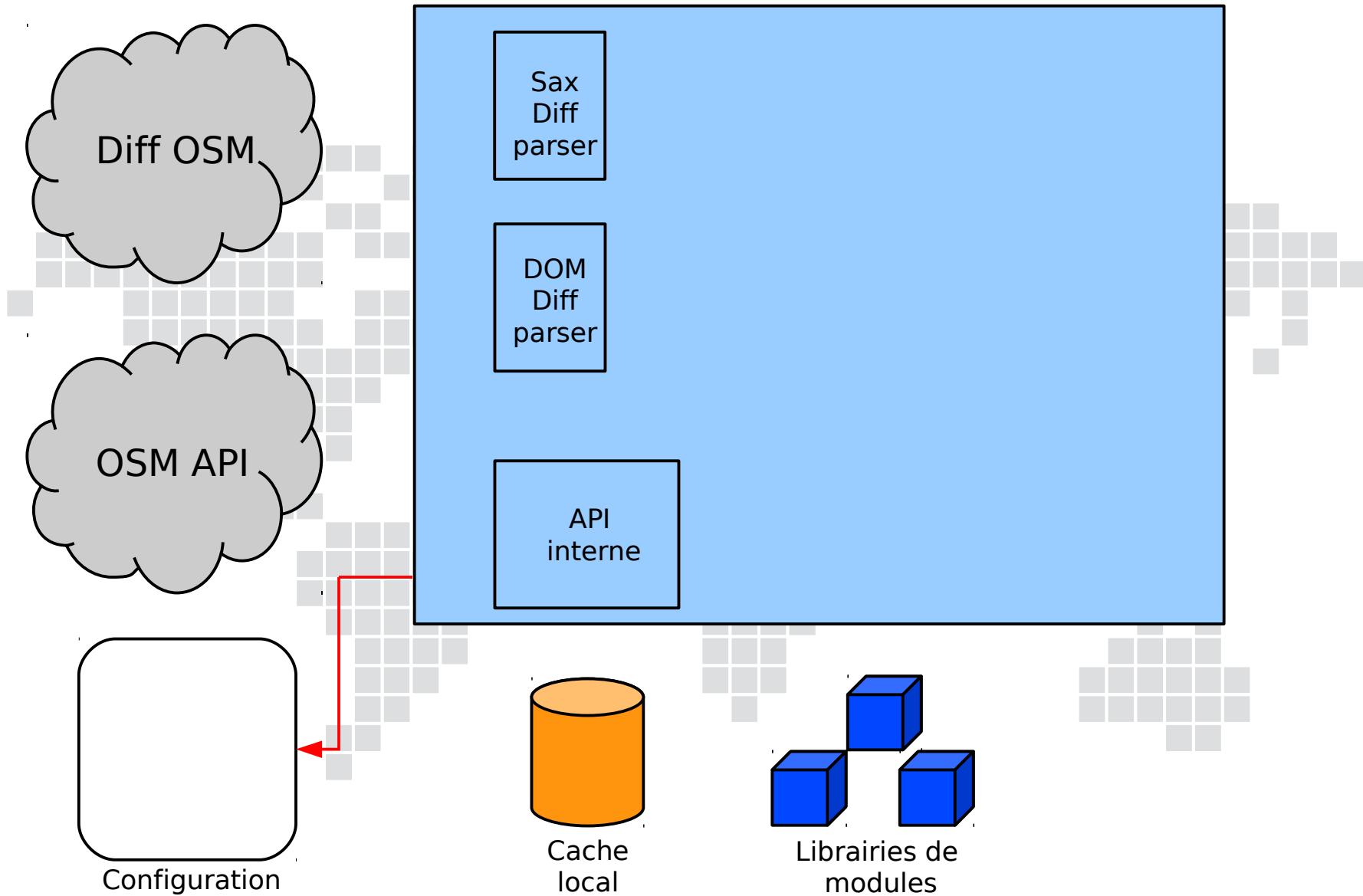
- Gestion de configuration
- Démarrage/arrêt de l analyse
- Rapports des modules
 - Format HTML



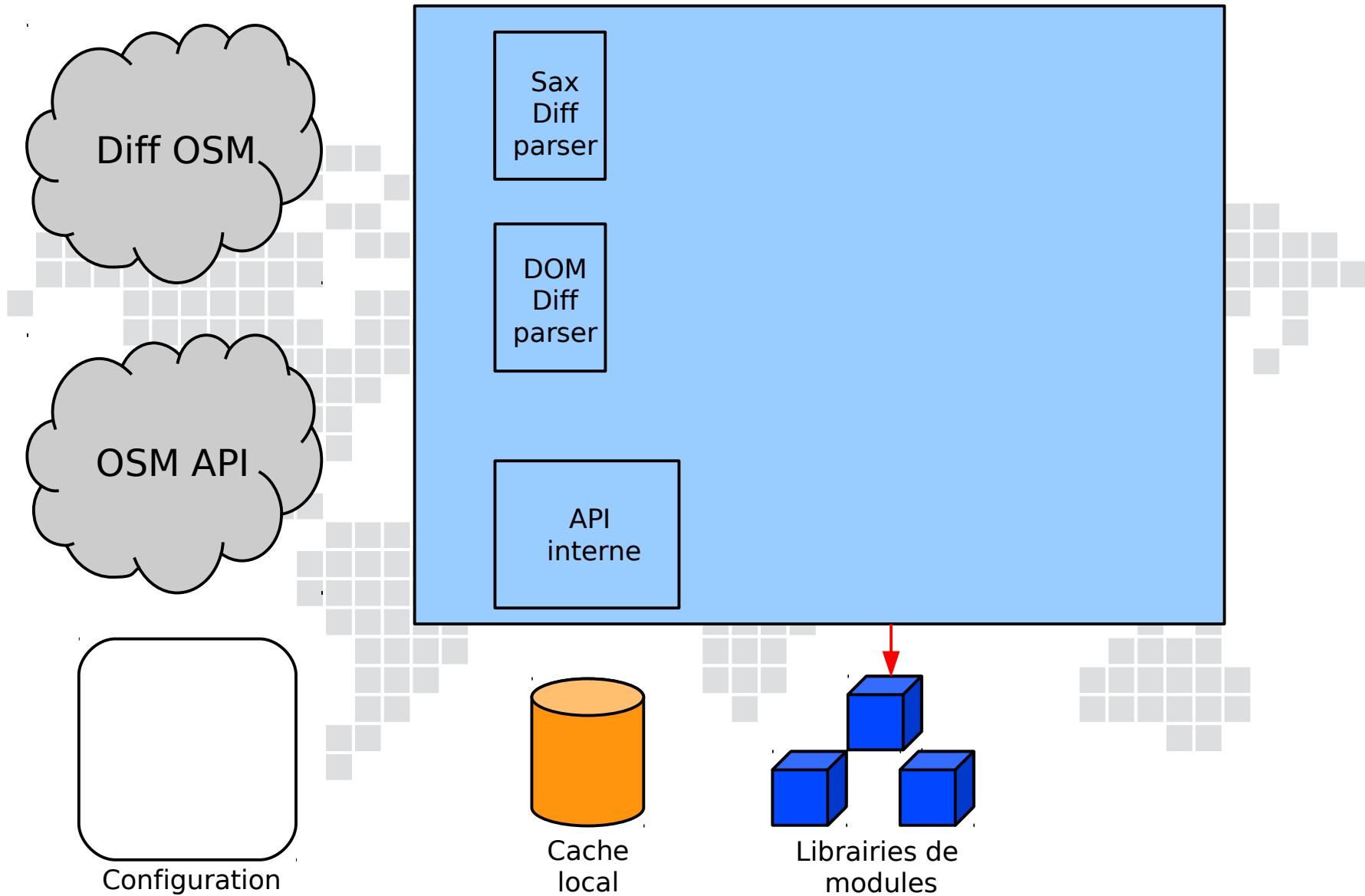
• En ligne de commande

- Fichier configuration
- Rapports des modules dans des fichiers

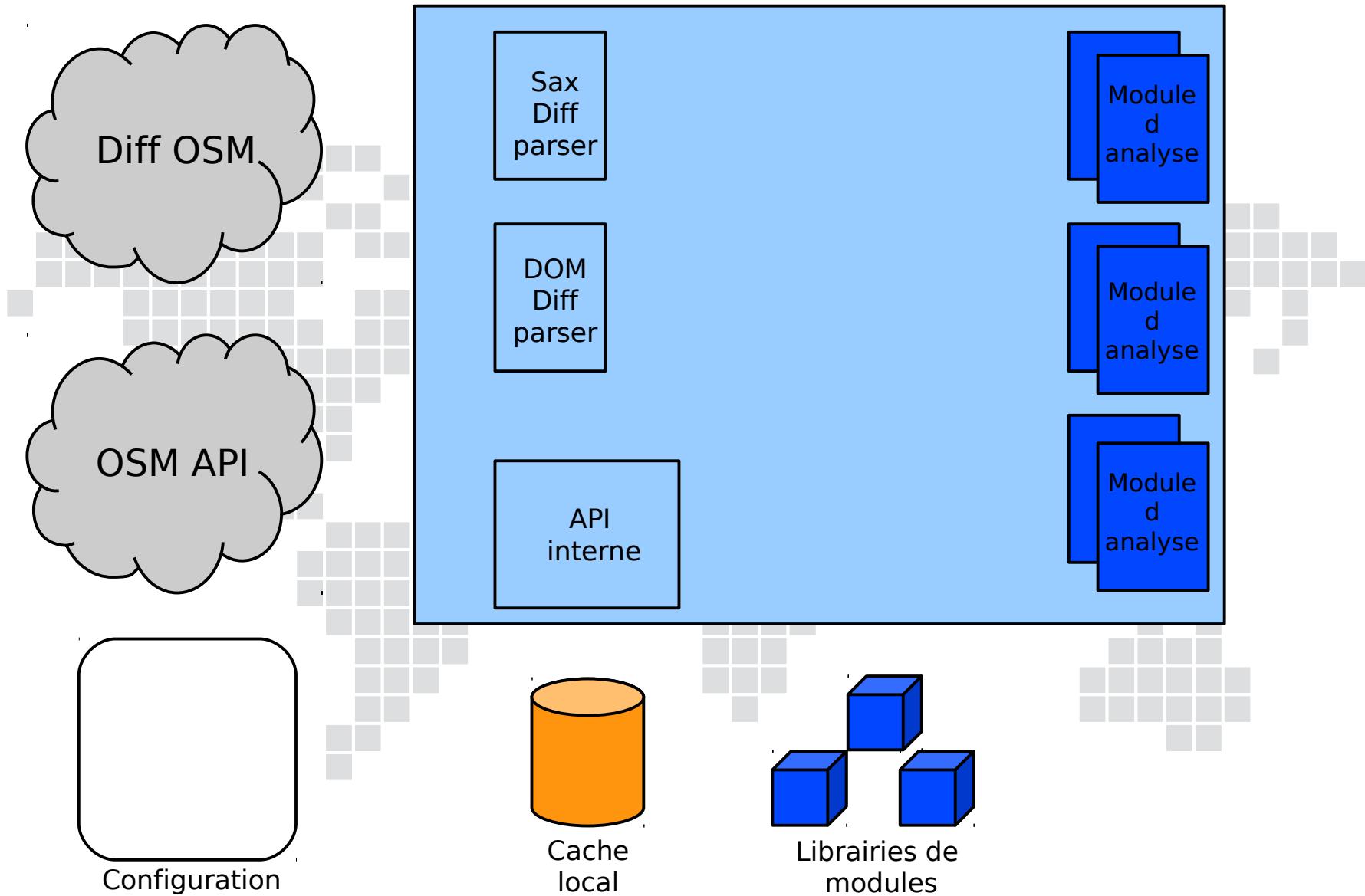
Principe de fonctionnement



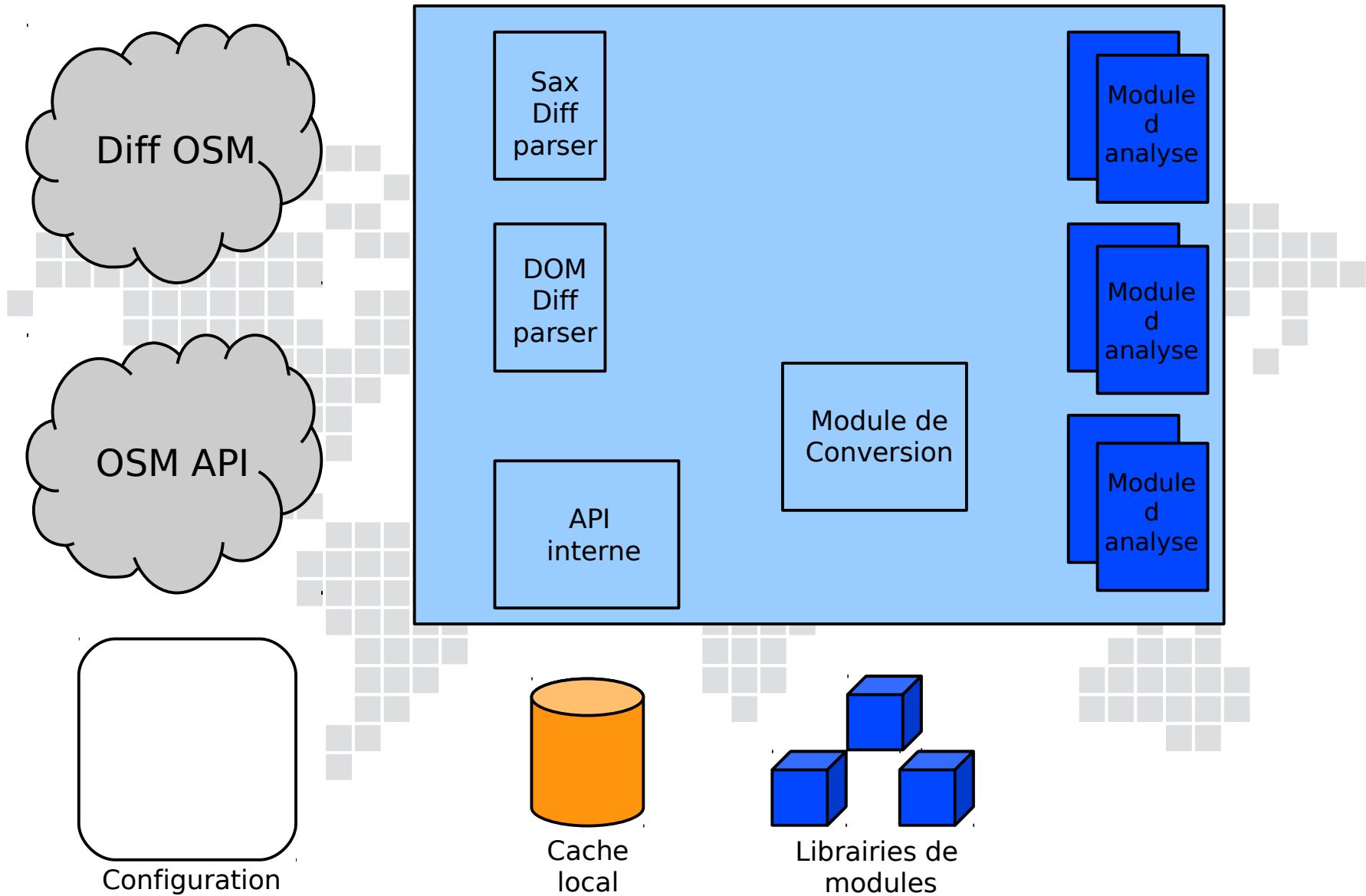
Principe de fonctionnement



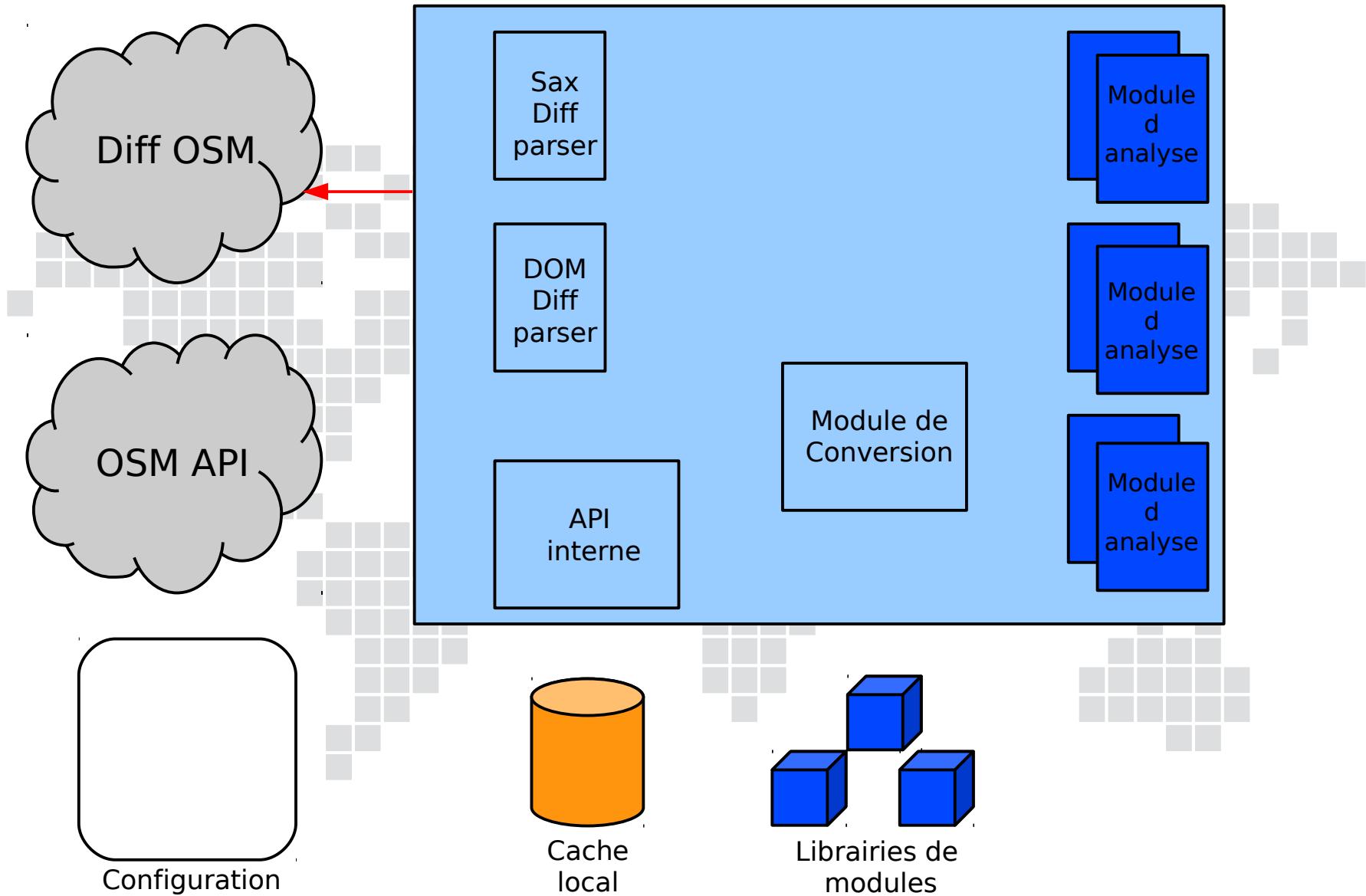
Principe de fonctionnement



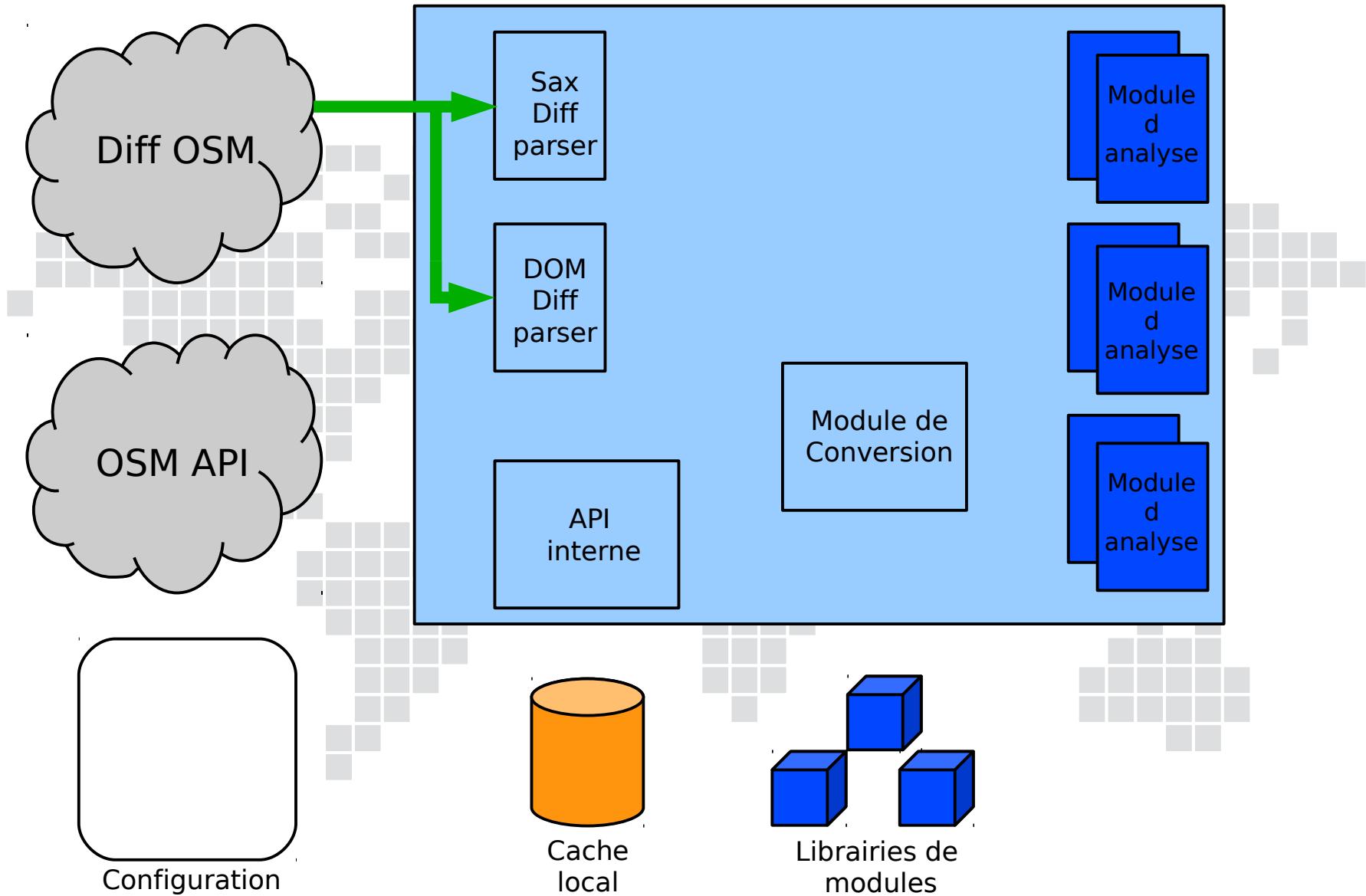
Principe de fonctionnement



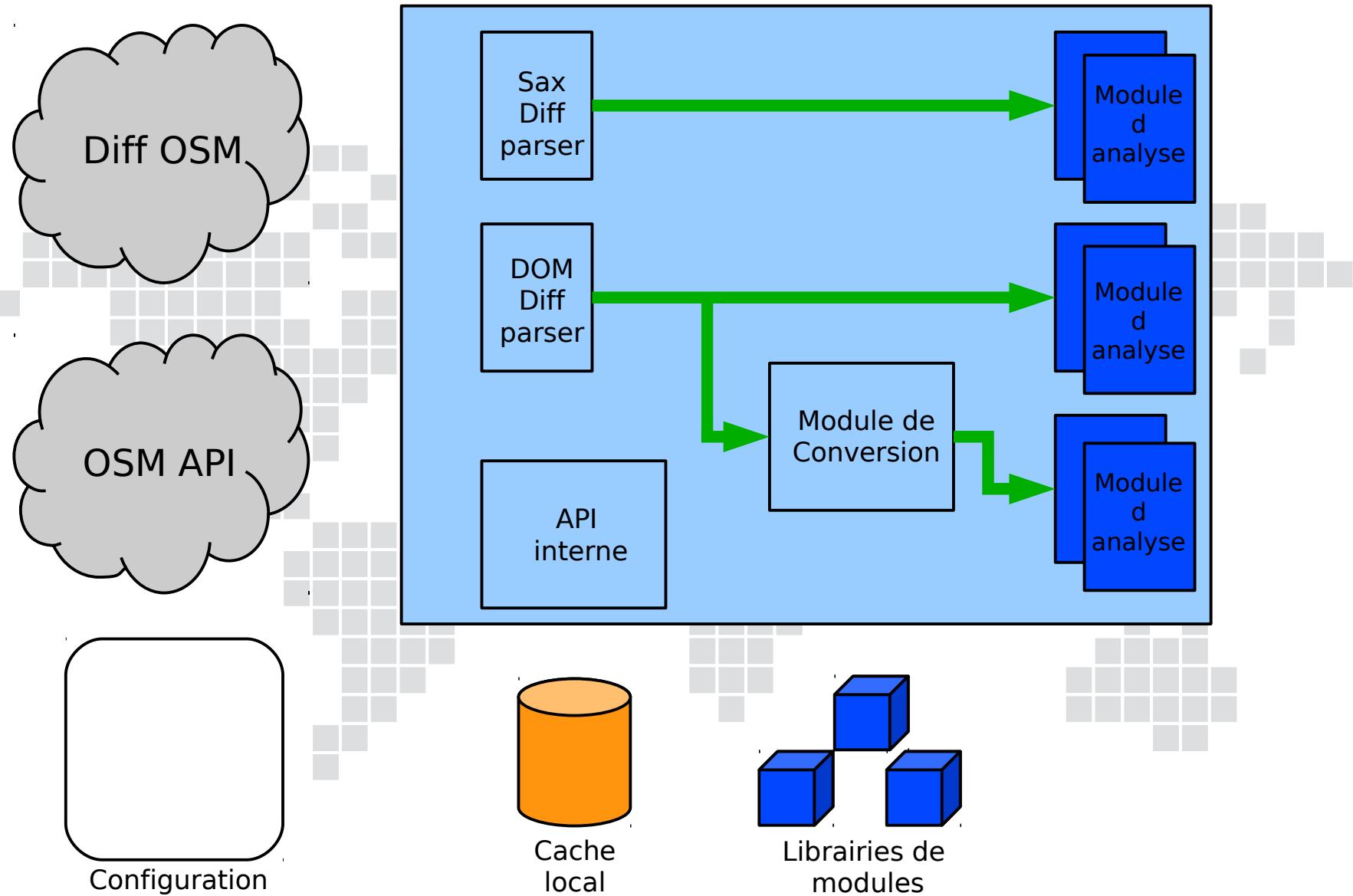
Principe de fonctionnement



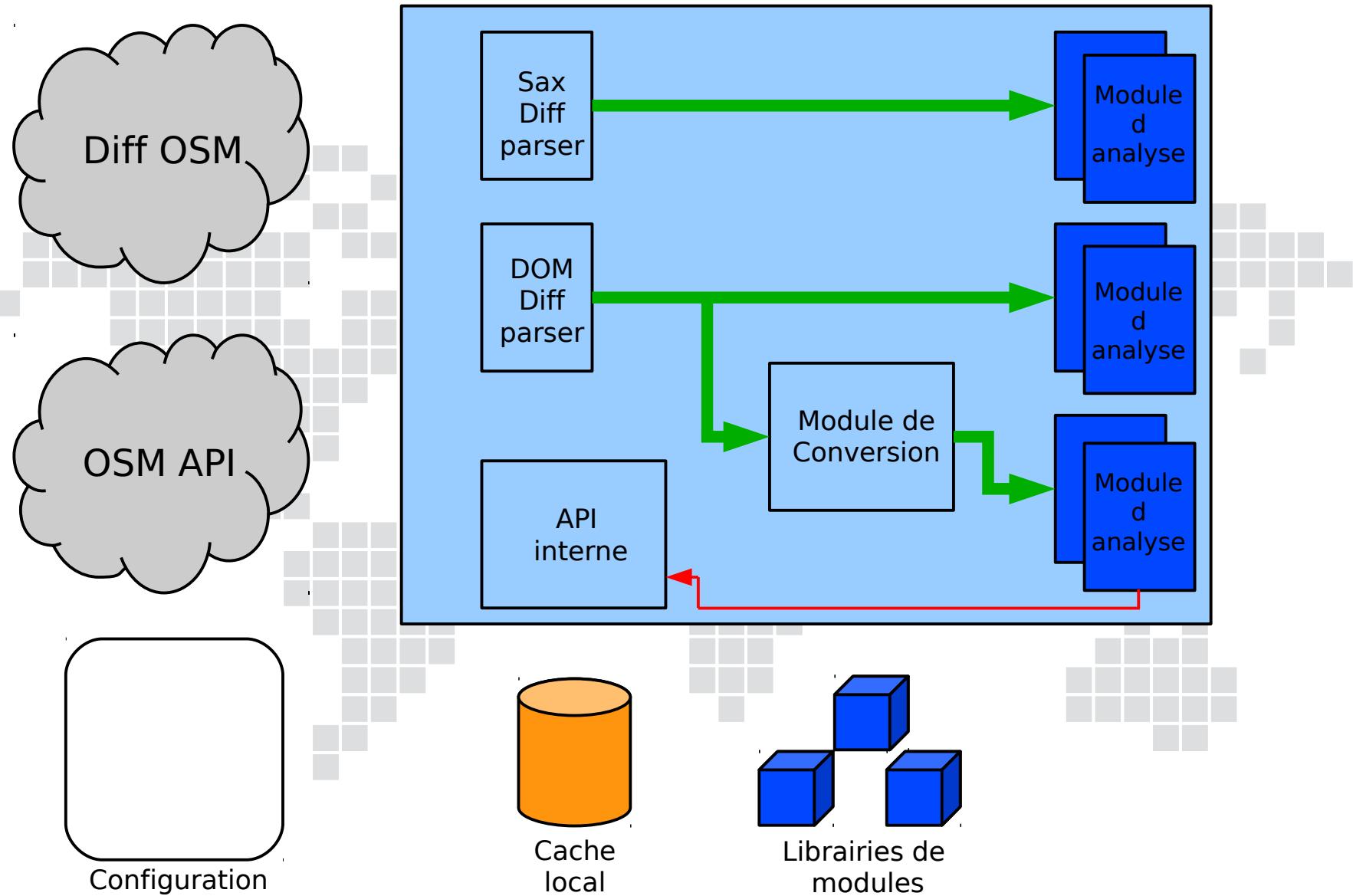
Principe de fonctionnement



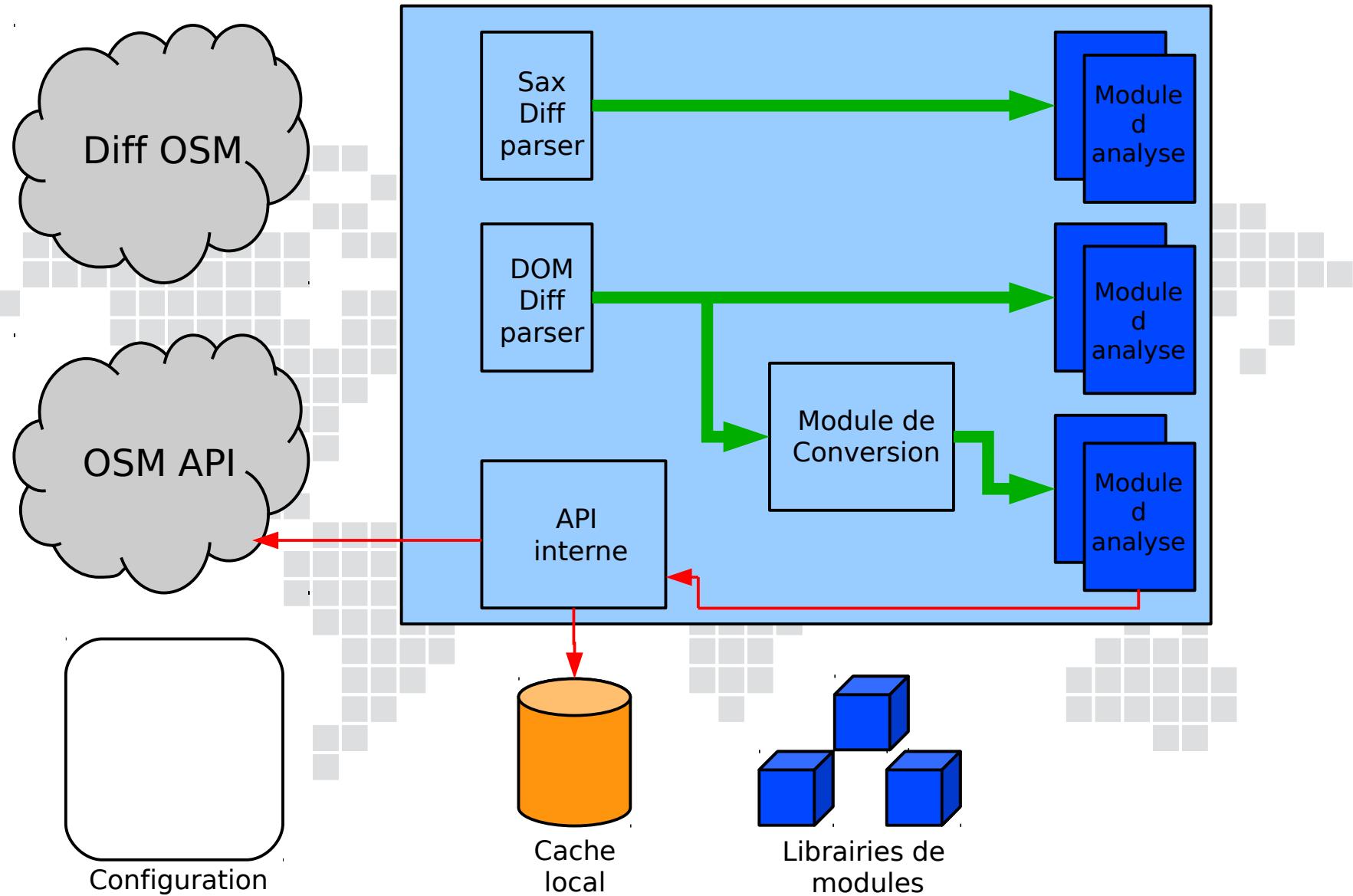
Principe de fonctionnement



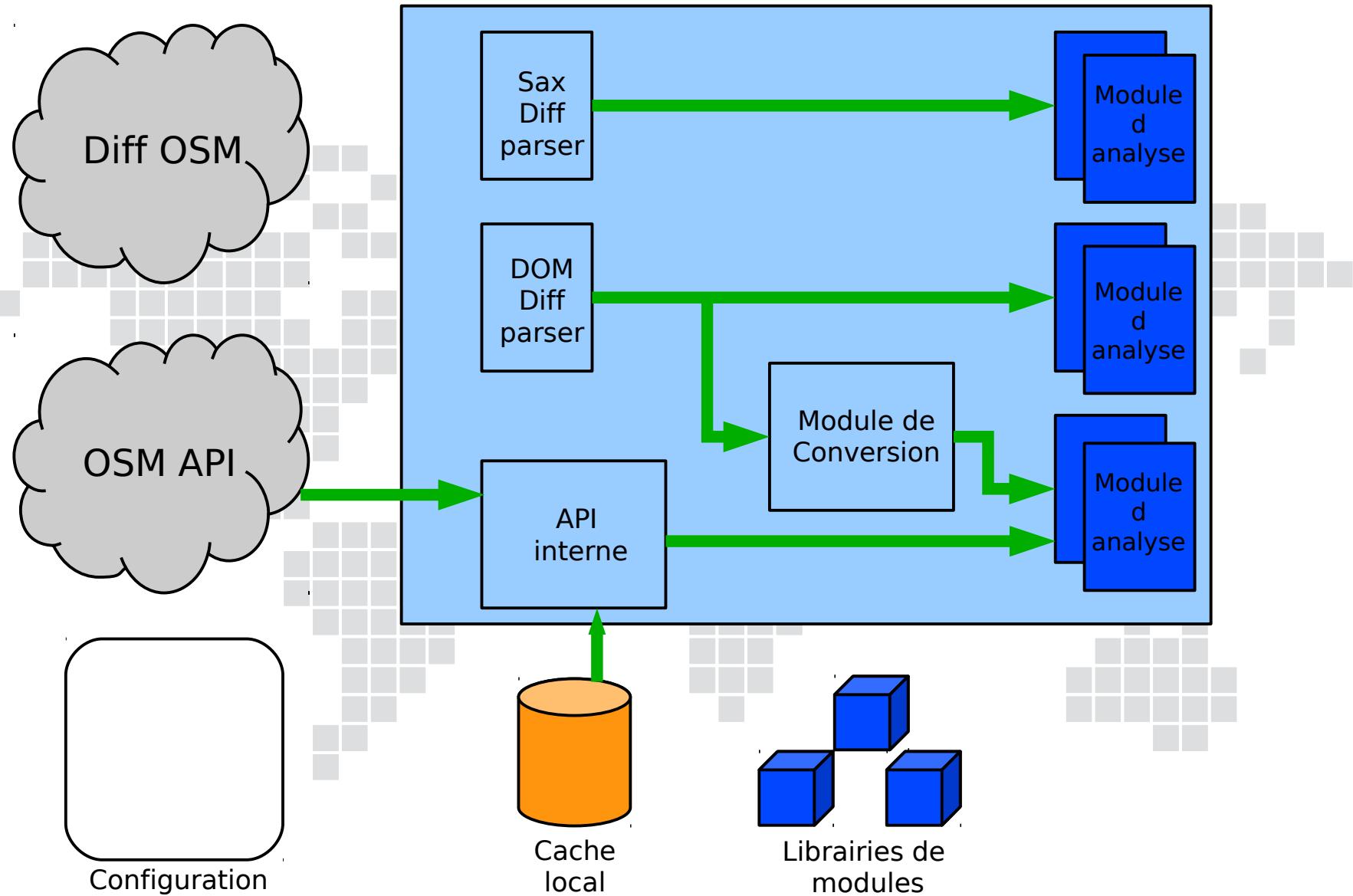
Principe de fonctionnement



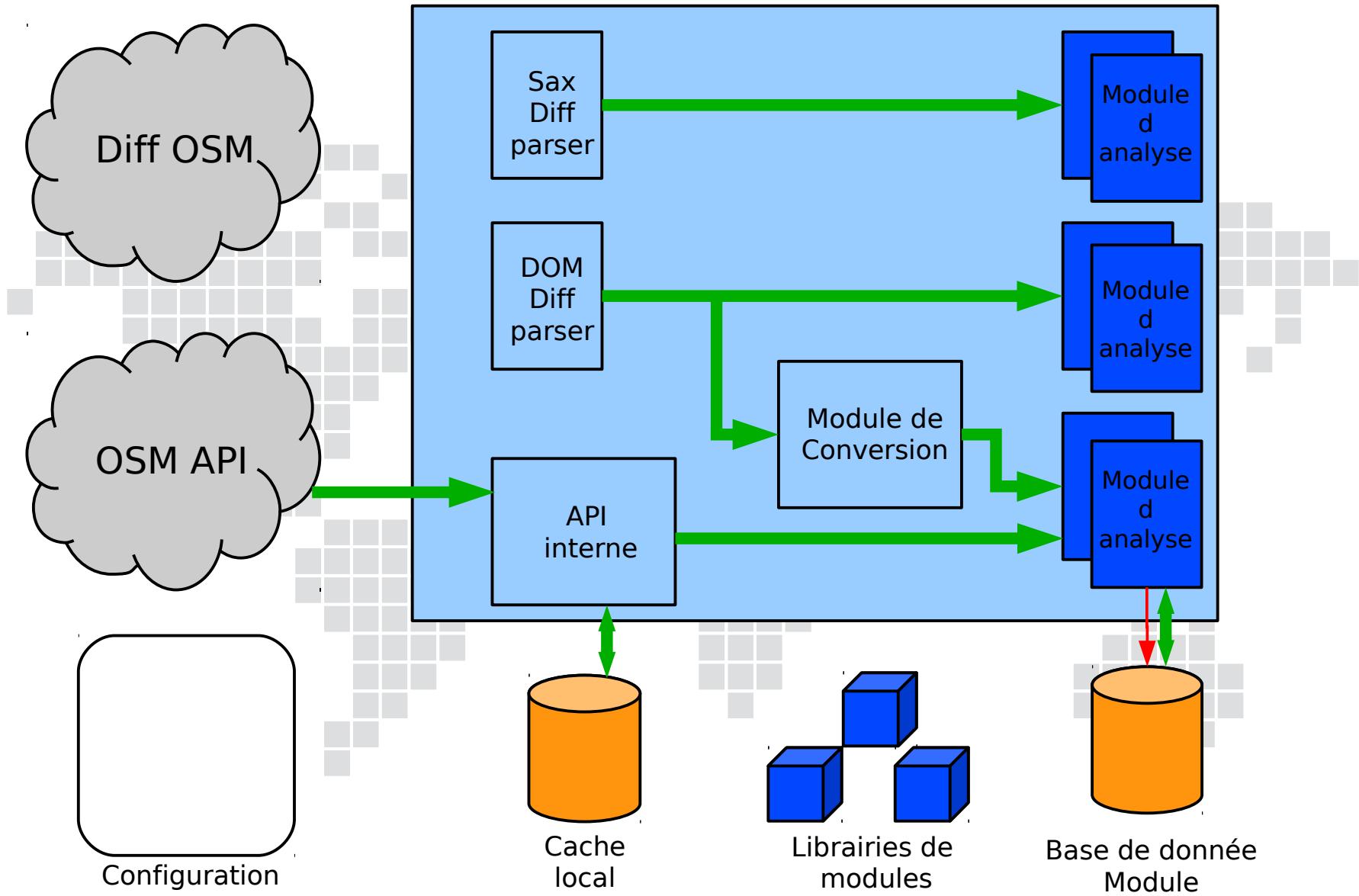
Principe de fonctionnement



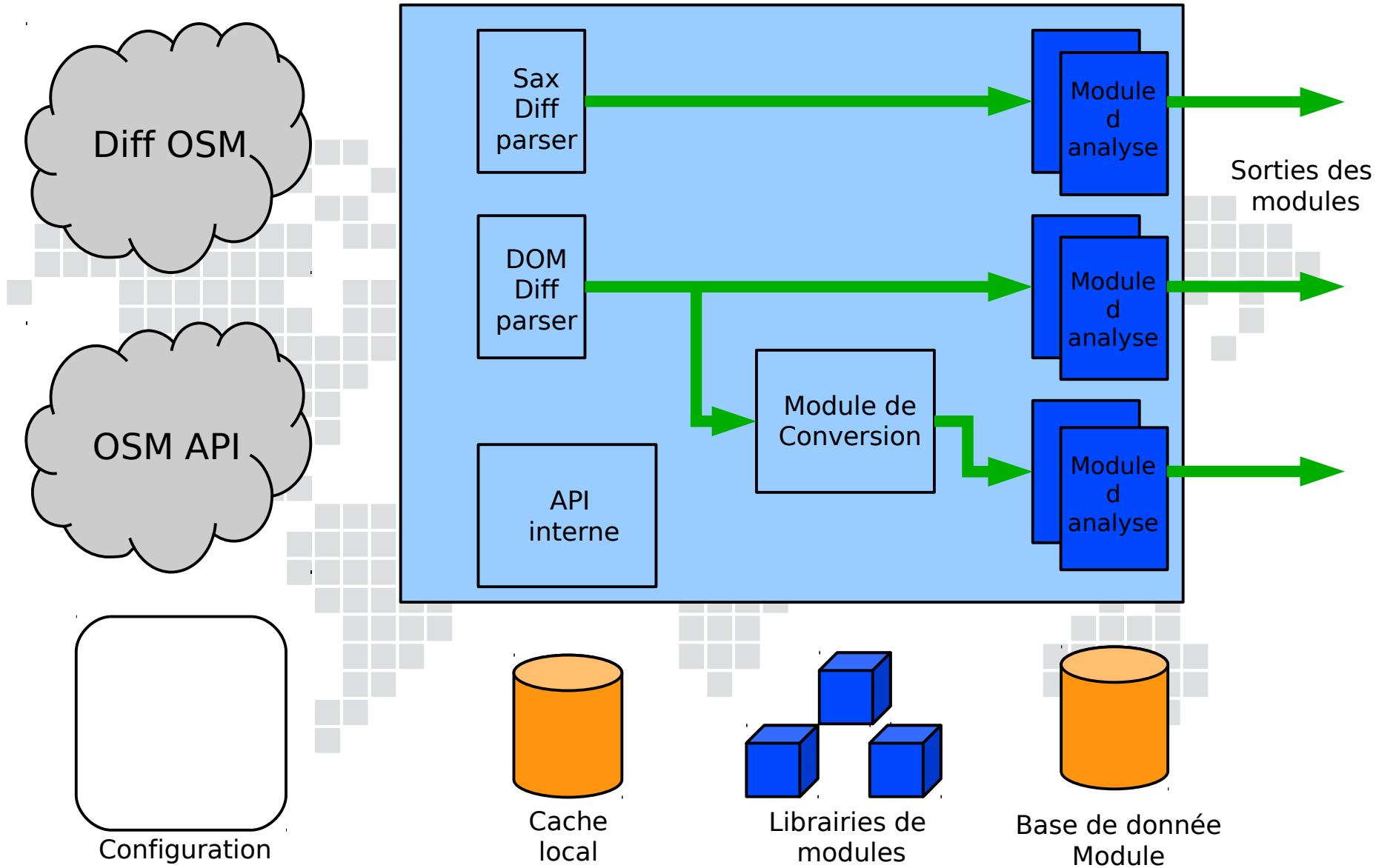
Principe de fonctionnement



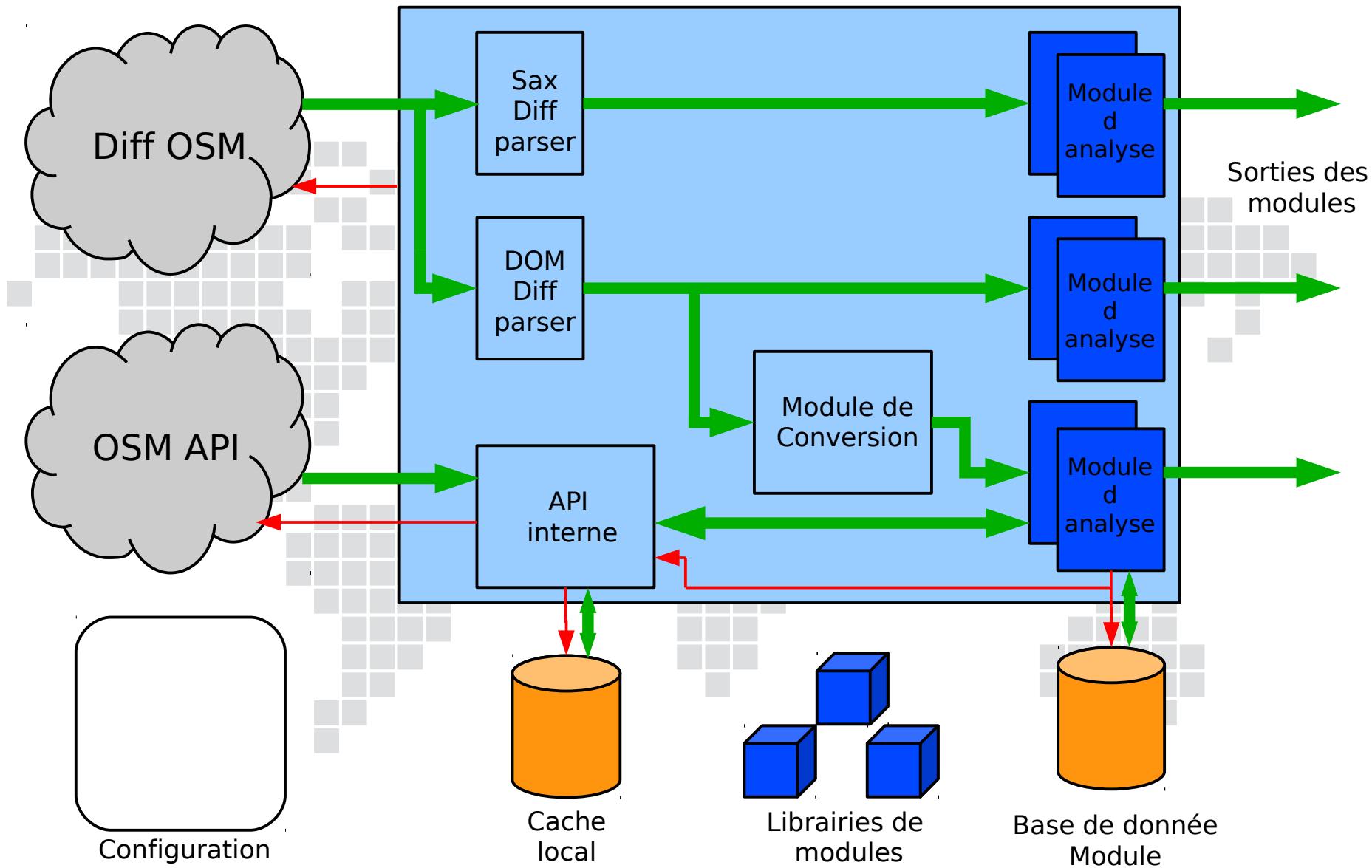
Principe de fonctionnement



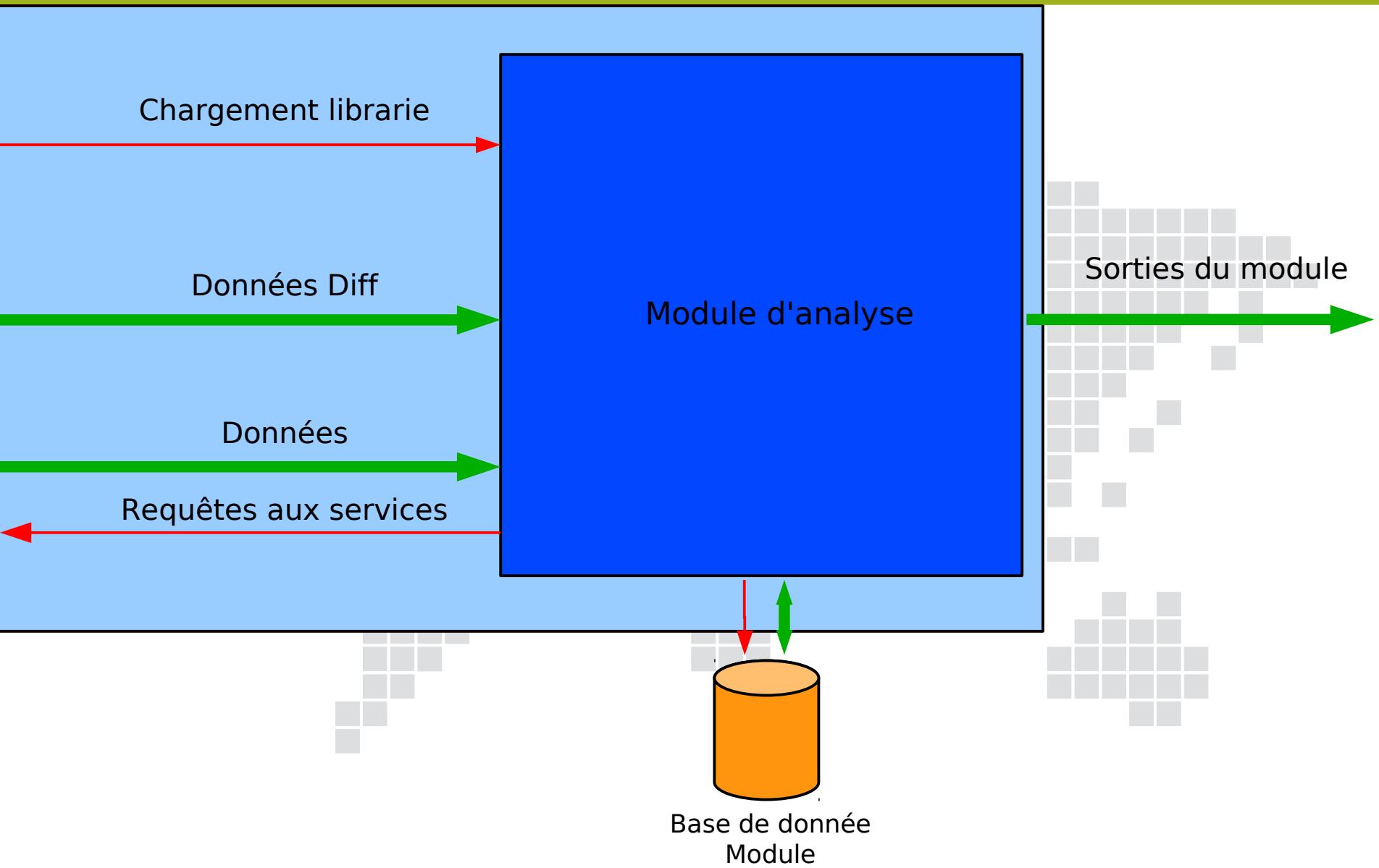
Principe de fonctionnement



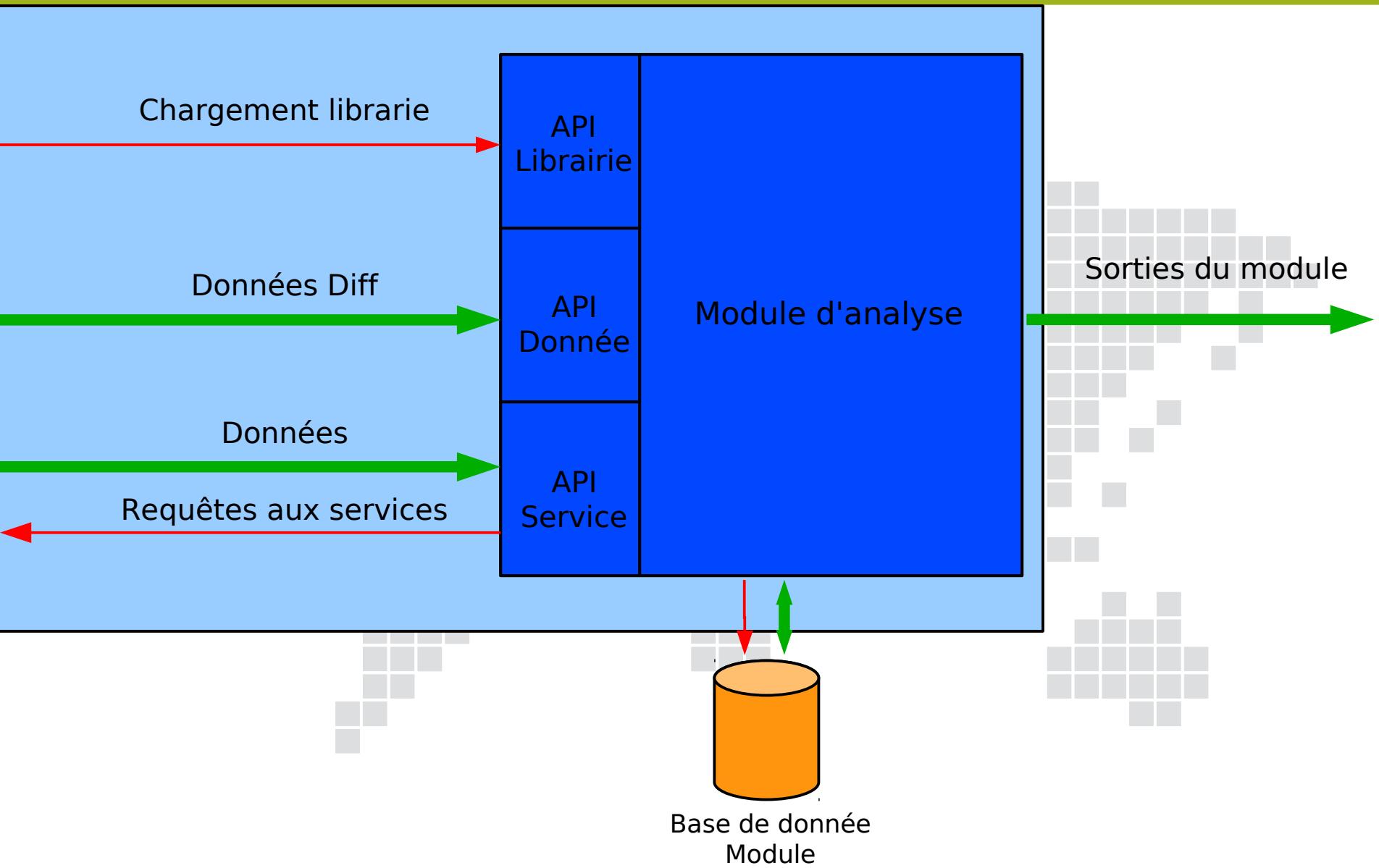
Principe de fonctionnement



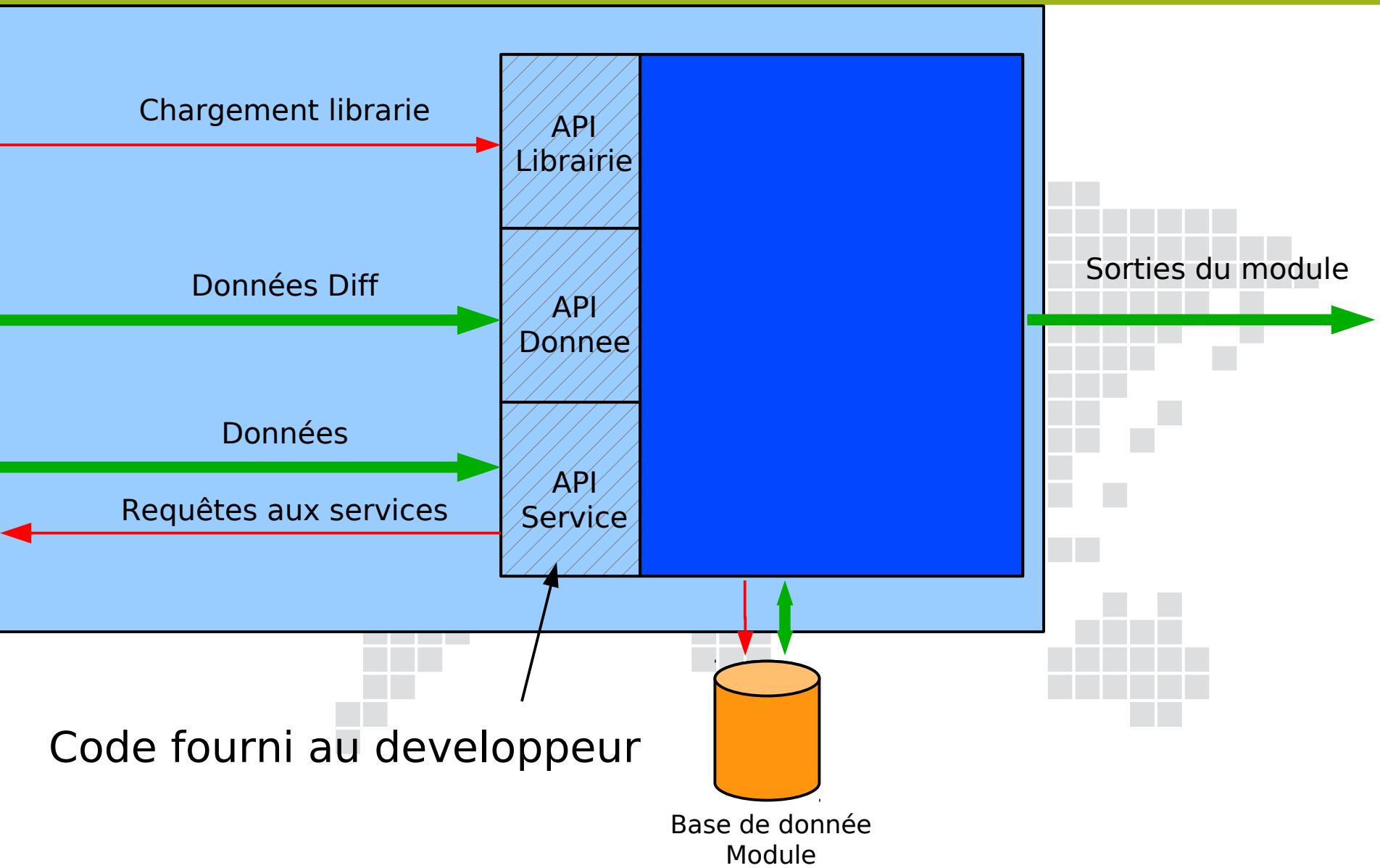
Architecture d'un module



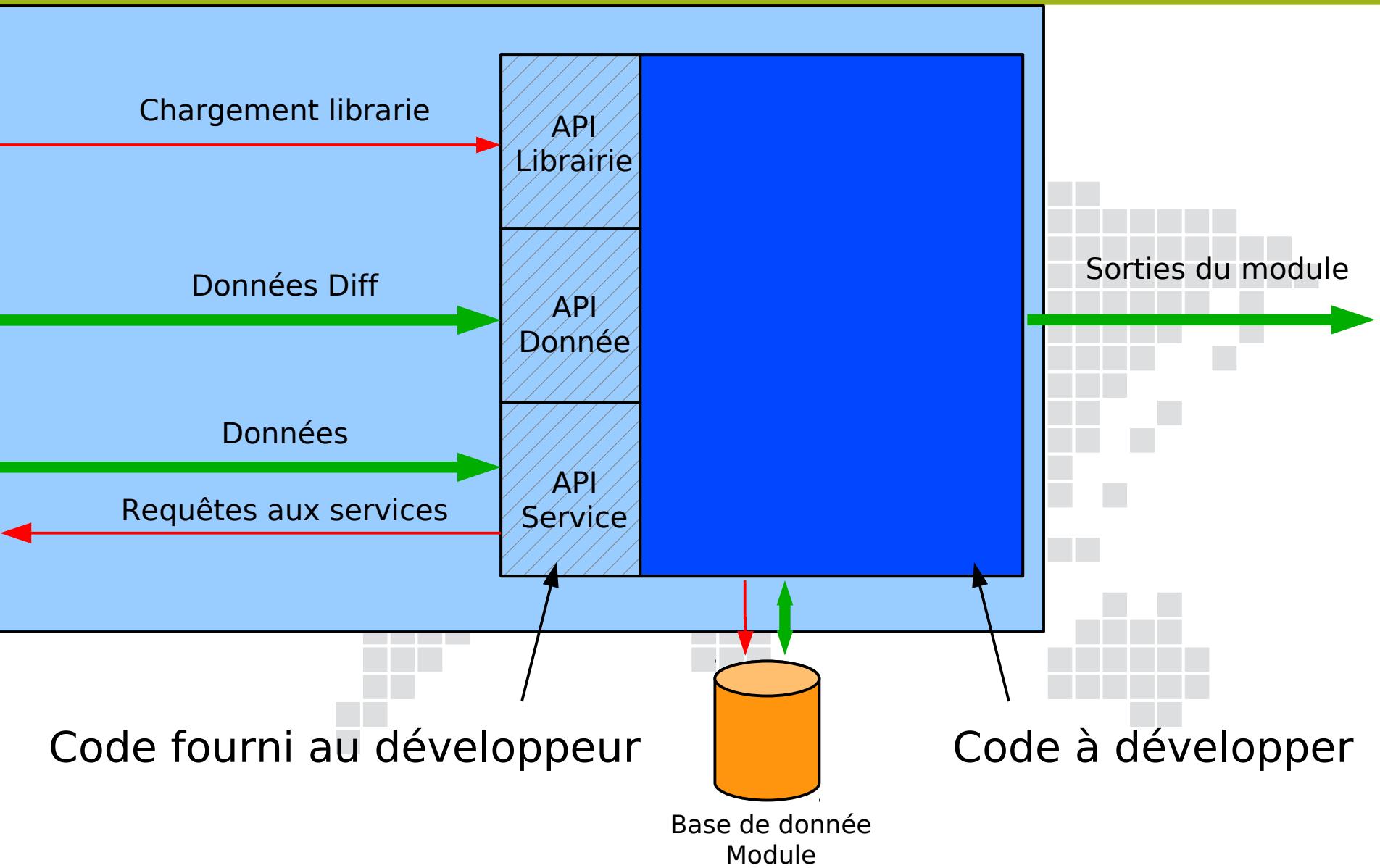
Architecture d'un module



Architecture d'un module



Architecture d'un module



- Défini la mise en forme des données Diff

- SAX : API orientée XML
 - Accès brutes limités, peu gourmand en mémoire
- DOM : API orientée XML
 - Accès brutes complets, gourmand en mémoire
- DOM/CPP : API orientée OSM/C++
 - Accès complet dédié OSM, gourmand en mémoire

- A venir

- SAX/CPP
 - Accès complet dédié OSM, peu gourmand en mémoire
- API plus haut niveau
 - Diff amélioré

• Détection de nouveaux utilisateurs

- Inscription datant de moins d'un mois

➤ [Rapport HTML](#)

Contributors since less than one month

[Threha](#) : February 21, 2013
[Protony](#) : February 11, 2013
[geesona](#) : February 14, 2013
[Arschloch](#) : February 22, 2013
[Nico Yan Bernard](#) : February 22, 2013
[Enrain](#) : January 29, 2013
[niranjana](#) : February 18, 2013
[Wolle DD](#) : December 25, 2012
[plauderkoffer](#) : December 14, 2012
[javicarrera](#) : February 14, 2013
[fmi2012](#) : February 03, 2013
[Zylog](#) : February 07, 2013
[j512](#) : February 22, 2013
[e_petkov](#) : February 15, 2013
[kräckel](#) : November 28, 2012
[Emmanuel Ezeji](#) : January 21, 2013

- Détection de nouveaux utilisateurs

- API Sax

Réception des données

Demande date d'inscription

Comparaison

Demande URL contributeur

Ajout au rapport

```

//-
void new_user_analyzer::start_element(const std::string & p_name){
    m_uid = 0;
    m_user_name = "";
}
void new_user_analyzer::get_attribute(const std::string & p_name,
                                      const std::string & p_value){
    if(p_name == "uid") m_uid = strtoull(p_value.c_str(),NULL,10);
    if(p_name == "user") m_user_name = p_value;
}
/*...
m_api.get_user_subscription_date(m_uid,m_user_name,l_date);
/*
*/
if((l_duration->tm_year - 70 )== 0 && l_duration->tm_mon < 1)
{
    std::string l_user_url;
    m_api.get_user_browser_url(l_user_url,m_uid,m_user_name);

    m_report << "<A HREF=\"" << l_user_url << "\">" << m_user_name ;
    m_report << "</A>" << l_date << "<BR>" << std::endl ;
}
/*
*/

```

- Statistiques sur les clefs de tag contenant une chaîne particulière

- Chaîne paramétrable
- Demande de Black Myst
- Utilisation pour les “projets de la semaine”
- Rapport détaillé des modifications
- Rapport de synthèse avec statistiques

wikipedia tag survey report

Statistics on edits from 2013-01-13T00:00:00Z to 2013-01-13T00:00:00Z

- Object created : 71
- Objects removed : 8
- Objects modified : 11358
- Tags added : 11262
- Tags modified : 73
- Tags removed : 11202
- Involved people : 57

User rank

OSM Contributor	Nb edits
malenki	22452
moasth	16
happy5214	15
HolgerJeromin	11
ohligser	10
bundesrainer	7
imagico	5
Kalaallit Nunaat	5
Rom1	5
gottferdom	5

- Suivi des modifications sur les objets créés ou modifiés par un utilisateur

- Nom d'utilisateur paramétrable
 - Initialisation par changeset ou fichier osm
 - Prise en compte des nouveaux objets teintés
 - Utilisé par Jean Claude Repetto
- **Rapport détaillé et simplifié des actions effectuées**

Geogast object report

[node 274262518](#) has been modified in version 8 by user [wheelmap_visitor](#) in [changeset 13720658](#)

- Tag "wheelchair" with value "no" has been added

[node 293791437](#) has been modified in version 4 by user [AmiFritz](#) in [changeset 13722049](#)

- Coordinates has changed from (48.0880089,7.4123621) to (48.0880013,7.4123702)

[node 1860363981](#) has been deleted in version 4 by user [Blekota](#) in [changeset 13723005](#)

[way 26784109](#) has been modified in version 12 by user [AmiFritz](#) in [changeset 13722049](#)

- [Node 1993724157](#) has been added to [Way 26784109](#)

[way 187869924](#) has been modified in version 2 by user [Blekota](#) in [changeset 13723005](#)

- [Node 1993724134](#) has been added to [Way 187869924](#)
- [Node 1993724124](#) has been added to [Way 187869924](#)
- [Node 1866866356](#) has been removed from [Way 187869924](#)
- [Node 1860363953](#) has been removed from [Way 187869924](#)

[relation 337767](#) has been modified in version 3 by user [Fred_OSM](#) in [changeset 13723019](#)

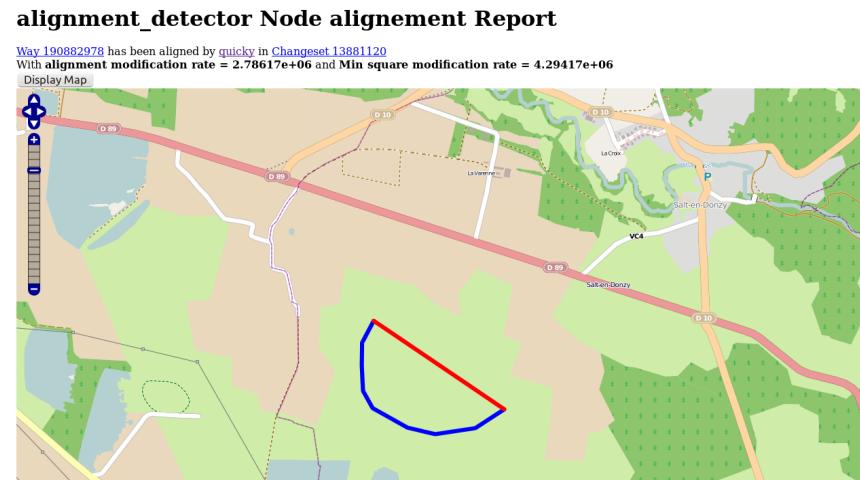
- [way 188751380](#) has been added with role "inner"
- [way 188751381](#) has been added with role "inner"

[relation 1923087](#) has been modified in version 9 by user [poppei82](#) in [changeset 13723009](#)

- [way 188751400](#) has been added with role "inner"

- Détection des ways réalignés

- Régression linéaire
- Comparaison avec la version précédente
- Alerte si différence supérieure au seuil paramétré
- Rapport avec la forme des deux versions de way sur fond OSM



- Récuperation de toutes les données teintées par un utilisateur

- Nom d'utilisateur paramétrable
- Recherche à partir de tous les changesets utilisateurs
- Utilisé pour CEDRIC007
- Fichiers .osm distinguant les données créées/modifiées/supprimées

- En projet

- Détection des cassures de limite administrative
- Détection des imports cadastres brutes
- Détection des suppressions massives
- Détection des sources non autorisées
- Suivi des bornes géodésiques
- Stats sur le temps écoulé entre chaque contribution

- Windows

- Rapide
 - Téléchargement des binaires
- Développeur
 - Installation de MinGW, QT
 - Téléchargement des sources
 - Compilation

- Linux

- Installation des paquets nécessaires à la compilation
- Téléchargement des sources
- Compilation

- Développer d'autres analyses
- Améliorer l'interface
- Ajouter des services
- Améliorer le cache
- Support du multi-coeur
- GUI de modules personnalisées
- Documenter

- Présentation sous licence CC-BY-SA



- Vous êtes libres de

- Partager : reproduire, distribuer et communiquer la présentation
- Remixer : adapter la présentation
- Utiliser cette présentation à des fins commerciales

- À condition de

- Indiquer les auteurs de la présentation
 - [Julien Thevenon](#)
- La partager sous des conditions identiques

- Présentation inspirée de

- Florian Lainéz
- Frederic Rodrigo
- Nicolas Moyroud



Merci de votre attention



A vos analyses!

Principe de fonctionnement

